

# Joint Relay Scheduling, Channel Access, and Power Allocation for Green Cognitive Radio Communications

Changqing Luo, Geyong Min, F. Richard Yu, Yan Zhang, Laurence T. Yang, and Victor C. M. Leung

## Abstract

The capacity of cognitive radio (CR) systems can be enhanced significantly by deploying relay nodes to exploit the spatial diversity. However, the inevitable imperfect sensing in CR has vital effects on the policy of relay selection, channel access, and power allocation that play key roles in the system capacity. The increase in transmission power can improve the system capacity, but results in high energy consumption, which incurs the increase of carbon emission and network operational cost. Most of the existing schemes on CR systems have not jointly considered the imperfect sensing scenario and the tradeoff between the system capacity and energy consumption. To fill in this gap, this paper proposes an energy-aware centralized relay selection scheme that takes into account the relay selection, channel access, and power allocation jointly in CR with imperfect sensing. Specifically, the CR system is formulated as a partially observable Markov decision process (POMDP) to achieve the goal of balancing the system capacity and energy consumption as well as maximizing the system reward. The optimal policy for relay selection, channel access, and power allocation is then derived by virtue of a dynamic programming approach. Furthermore, a dimension reduction strategy is applied to reduce its high computation complexity and the upper bound is further discussed to show the performance loss of the proposed scheme. Extensive simulation experiments and results are presented and analysed to demonstrate the significant performance improvement compared to the existing schemes. The results show that the received reward increases more than 50% and the network lifetime increases more than 35%, but the system capacity is reduced less than 6% only.

C. Luo and L. T. Yang are with the School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, China (emails: [chqluo2013@gmail.com](mailto:chqluo2013@gmail.com) and [ltyang@gmail.com](mailto:ltyang@gmail.com)); L. T. Yang is also with the Department of Computer Science, St. Francis Xavier University, Antigonish, Canada.

G. Min is with the College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, EX4 4QF, United Kingdom (email: [g.min@exeter.ac.uk](mailto:g.min@exeter.ac.uk)).

F. R. Yu is with the Department of Systems and Computer Engineering, Carleton University, Ottawa, Canada (email: [Richard\\_yu@carleton.ca](mailto:Richard_yu@carleton.ca)).

Y. Zhang is with the Simula Research Laboratory, 1325 Lysaker, Norway (email: [yanzhang@ieee.org](mailto:yanzhang@ieee.org)).

V. C. M. Leung is with the Department of Electrical and Computer Engineering, the University of British Columbia, Vancouver, BC V6T 1Z4 Canada (email: [vleung@ece.ubc.ca](mailto:vleung@ece.ubc.ca)).

### Index Terms

Cognitive Radio; Imperfect Sensing; Relay Selection; Channel Access; Energy Efficiency; Green Communications.

## I. INTRODUCTION

Due to the fact that the spectrum band is getting crowded for the emergence of new mobile applications whilst an enormous amount of spectrum is still under-utilized, the ever-increasing demand for higher capacity in wireless communications has motivated the technology of cognitive radio (CR). As an effective approach to improving the utilization of scarce spectrum, the CR technology is the key to realize the dynamic spectrum access (DSA) where secondary users can opportunistically utilize the unused licensed spectrum [1]. Because the vacant spectrum is opportunistically used by secondary users through exploiting advanced spectrum sensing technology [2], the spectrum efficiency can be improved significantly and the shortage of radio resource can be alleviated as well [3].

The distinct advantage of CR is to improve the system performance through dynamically adapting the operating parameters according to the surrounding radio environment [4]. In recent years, many research efforts have been made to improve the performance of CR. For example, Bansal, Hossain, and Bhargava [5] proposed a power allocation scheme to maximize the system capacity. The Medium Access Control (MAC) protocols in CR networks were thoroughly studied in [6]. Luo *et al.* [7] addressed the TCP performance improvement problem in CR networks from a perspective of cross-layer design. In particular, the spectrum sensing, channel access, and configuration of operating parameters in physical/data-link layers were jointly investigated to optimize the TCP performance. Fu, Zhang, and Huang [8] investigated energy-efficient transmissions for Multiple-Input-Multiple-Output (MIMO) cognitive radio networks.

On the other hand, cooperative communication (CC) has been identified as a promising technology to improve the system performance significantly by exploiting the spatial diversity [9]. In CC,

an intermediate node with a single antenna relays the communication between a source node and a destination, and thus a virtual MIMO system is formed to achieve the improved communication performance. Many cooperation techniques (such as decode-and-forward (DF), amplify-and-forward (AF), and coded cooperation) have been proposed [10]. Furthermore, numerous studies have been conducted to improve the system performance through the use of CC in CR systems [11], [12], [13], [14]. For instance, Li *et al.* [15] investigated the problem of joint relay selection and power allocation in CR to maximize the system throughput and a dual method based optimal approach was proposed to address the system throughput optimization problem. Liu *et al.* [16] proposed a cooperative beamforming scheme to forward messages in busy time slots from a perspective of cross-layer design so as to attain the cooperative diversity gain and improve the system performance. Li *et al.* [17] investigated the capacity maximization problem in cooperative CR systems where the relay assignment and channel allocation are jointly considered. Luo *et al.* [18] considered the energy efficiency problem in cognitive radio cooperative communications under the spectrum underlay paradigm.

However, the work on joint relay selection, channel access, and power allocation in CR systems with imperfect sensing is still in its infancy. Actually, the practical CR sensors are bound to sensing errors and the imperfect sensing is inevitable due to the hardware limitation [19]. Therefore, it is very necessary to consider imperfect sensing in CR because the system performance can be affected directly by the imperfect knowledge of the spectrum state. Furthermore, the reduction in energy consumption in wireless networks is stringent [20], [21], and thus energy consumption should be also considered in CR systems. It is a common sense that reduction in energy consumption may result in the degradation of system capacity. Therefore, the system capacity and energy consumption should be simultaneously investigated and balanced for the efficient design and implementation of CR systems.

To this end, this paper proposes an energy-aware joint relay selection, channel access, and

power allocation scheme for green CR systems, which can balance the system capacity and the energy consumption. The major contributions of this paper are summarized as follows.

- 1) The CR system with imperfect sensing is formulated as a partially observable Markov decision process (POMDP) where the most likely system state can be derived through the information state that encaptures the history of system state and decision.
- 2) The optimal policy for relay selection, channel access, and power allocation is obtained by solving the POMDP formulation and the solution is derived through a dynamic programming approach. However, the computation complexity is very high since it is highly relative to the number of system states. A strategy for reducing dimension is then applied to decrease the computation complexity dramatically. The upper bound is investigated to evaluate the performance loss of the proposed scheme.
- 3) The system capacity and the energy consumption are concurrently taken into account and their tradeoff is achieved in this proposed scheme. The transmission power is adjusted according to the tradeoff so as to achieve green CR communications.
- 4) Extensive simulation experiments and results are presented and analysed to investigate the performance of the proposed scheme and evaluate the effects of the false alarm probability of the spectrum sensing, radio channel estimation error, and the dynamic behaviors of primary users. The results show that the system reward can be improved more than 50% compared to the existing schemes and the network lifetime increases more than 35% while the system capacity declines slightly.

The rest of this paper is organized as follows. Section II presents the system scenario, the CC with AF relaying, and the energy consumption model as well as the finite state Markov channel model. In Section III, CR system with imperfect sensing is formulated as a POMDP system and the optimal policy is then obtained through the dynamic programming and the dimension reduction strategy. Extensive simulation results are presented and analyzed for performance

evaluation in Section IV. Finally, Section V concludes this paper.

## II. SYSTEM DESCRIPTION

As shown in Fig. 1, an overlay CR system coexisting with a primary network is considered. Both primary and secondary users share a block of spectrum consisting of  $L$  radio channels, each of which has bandwidth  $\mathcal{W}(l)$ ,  $1 \leq l \leq L$ . In the primary network, a primary transmitter ( $\mathbf{P}_T$ ) communicates with its corresponding receiver ( $\mathbf{P}_R$ ) via a base station over radio channels. Meanwhile, a secondary source ( $\mathbf{S}$ ) transmits information towards a secondary destination ( $\mathbf{D}$ ) with the help of a mobile user who works as a relay ( $\mathbf{R}$ ). All mobile user that can overhear the information from the source node consist of a potential relay set  $R = \{R_n | n \in \mathcal{N} = \{1, 2, \dots, N\}\}$ .  $h_{S,D}$ ,  $\{h_{S,R_n} | n \in \mathcal{N}\}$ , and  $\{h_{R_n,D} | n \in \mathcal{N}\}$  are denoted as the channel gain that captures the effects of path-loss, shadowing, and fading with the respective channel of Source/Destination ( $\mathbf{S-D}$ ) link, Source/Relay ( $\mathbf{S-R}$ ) link, and Relay/Destination ( $\mathbf{R-D}$ ) link, respectively. Typically, the channel gain is affected by several factors, such as the distance between the two nodes connecting to a channel, the shadowing, and the mobility of the node in wireless networks. Moreover, this paper considers a finite state Markov channel (FSMC) model that has been widely accepted as an effective approach to characterizing the structure of the block fading process [22]. There are several channel models that suffer block fading process, like Rayleigh block-fading channel model [22] and Nakagami block-fading channel model [23]. Here, a Rayleigh block-fading channel model is considered and all users experience such fading. In addition, each node in this scenario works in a half-duplex mode, and thus cannot transmit and receive information simultaneously so as to avoid self-interfering. Finally, a time-slot system is also considered, where the time is divided into slots with equal length  $T$  and slot  $k$  is referred to as the discrete time period  $[kT, (k+1)T]$ . The duration of one time slot is defined as the time interval between two continuous transmissions.

### A. CC with AF Relaying

CC can significantly improve the system capacity or other performance metrics by exploiting the radio broadcasting and relaying. The typical cooperative model is the three-node CC including the source, destination, and relay. The relay forwards the information overheard from the source to the destination in a desired way of relaying. In this paper, an amplify-and-forward (AF) coding cooperation is considered as a relaying protocol.

Under this mode, as shown in Fig. 1, **S** broadcasts its information to its intended **D** and meanwhile **R** also overhears the information due to the nature of broadcasting. Then, **R** will amplify and forward the received signals to **D** [9]. Based on the results of classic three-node CC with AF relaying, the achievable system capacity for a given relay  $R_n$  with the accessed channel  $l$ , the achievable system capacity,  $C_{AF}$ , is obtained by [9]

$$C_{AF}(R_n, l, P_R^{tx}) = \mathcal{W}(l) \cdot \log_2 \left( 1 + SNR_{S,D} + \frac{SNR_{S,R} \cdot SNR_{R,D}}{SNR_{S,R} + SNR_{R,D} + 1} \right), \quad (1)$$

where  $SNR_{S,D} = \frac{P_S^{tx}}{\sigma^2} |h_{S,D}|^2$ ,  $SNR_{S,R_n} = \frac{P_S^{tx}}{\sigma^2} |h_{S,R_n}|^2$ , and  $SNR_{R_n,D} = \frac{P_{R_n}^{tx}}{\sigma^2} |h_{R_n,D}|^2$ ;  $P_S^{tx}$  and  $P_{R_n}^{tx}$  are presented the transmission power at **S** and **R**, respectively;  $n_R$  and  $n_D$  represent the additive Gaussian white noise (AGWN) received by **R** and **D**. The noise power at these two nodes is denoted by  $\sigma^2$ . Eq. (1) reveals that the achievable system capacity is directly affected by the transmission power at **S** and **R** and the quality of link **S-R**, **S-D** and **R-D**. Hence, the relay selection, channel access, and transmission power should be carefully considered for efficient design and implementation of CR systems.

### B. Energy Consumption Model

In wireless networks, most wireless mobile devices are powered by batteries with limited energy. In order to prolong the network lifetime, the energy consumption should be considered as an important factor in the process of scheduling relay node, accessing channel, and allocating transmission power. In opportunistic spectrum access systems, a CR user firstly decides whether

to sense a channel, then detects the channel if the channel sensing decision is made, and finally accesses the channel to transmit information if the channel is available. The energy is accordingly consumed for each operation in the process. The energy consumption for channel sensing is denoted by  $P^{cs}$  and the energy consumption for information transmission is presented by  $P^{tx}$ . In practical systems, compared to the energy consumption for channel sensing and information transmission, the energy consumption for determining whether to sense a channel is negligible, and thus is not taken into account in the proposed scheme. Moreover, the sensing power is a constant during the sensing process and no power will be consumed when a node is in sleeping state. In addition, the transmission power  $P^{tx}$  is limited by the maximum transmission power  $P^{mtx}$  for the devices used by CR users.

Let  $t_{cs}$  and  $t_{tx}$  denote the time taken by sensing channel and the time for transmitting data, respectively. Their sum meets  $t_{cs} + t_{tx} = T$ . As a result, in time slot  $k$ , the energy consumption,  $E_{ec}(k)$ , is given by

$$E_{ec}(k) = \begin{cases} 0, & \text{if there is no sensing,} \\ P^{cs} \cdot t_{cs}, & \text{if a node senses a channel,} \\ P^{cs} \cdot t_{cs} + P^{tx} \cdot t_{tx}, & \text{if a node is used for transmission.} \end{cases} \quad (2)$$

In this paper, each intermediate node is powered by battery with the identical finite initial energy  $E_0$ . Let  $E(k)$  denote the residual energy of a CR user at the beginning of slot  $k$ . Accordingly, after a slot, the residual energy at the beginning of slot  $k + 1$  is

$$E(k + 1) = E(k) - E_{ec}(k). \quad (3)$$

### C. Opportunistic Channel Availability and Finite State Markov Channel Model

In this paper, a Rayleigh block-fading channel which can keep stable during the transmission of one data block is considered. In the FSMC model, the range of channel gain is partitioned into discrete levels, each of which corresponds to a state in the Markov chain. Therefore, an FSMC is characterized by a state set, i.e.,  $\mathcal{C} = \{1, 2, \dots, S\}$ , where  $|S|$  is the number of states.

Moreover, the busy state that a channel is in use by primary users should also be integrated into the FSMC model, and is represented by 0. Thus,  $\mathcal{C} = \{0, 1, 2, \dots, S\}$ . Given the knowledge of the block fading process and the dynamics of the primary user using spectrum, the channel state transition probability  $p(i, j)$  can be derived naturally.

### III. A SOLUTION TO THE RELAY SELECTION, CHANNEL ACCESS AND POWER ALLOCATION PROBLEM

Eqs. (1) and (2) presented in the previous section show that the achievable system capacity and the energy consumption for transmission are affected directly by three factors: 1) the relay selection that has significant influence on the channel quality of **S-R** and **R-D**; 2) the channel access that imposes the effect on the achievable system capacity; 3) the power allocation that affects the energy consumption for information transmission in each slot. Therefore, the relay selection, channel access and power allocation need to be jointly considered to improve the performance of CR systems.

This section presents a new stochastic optimization method for solving the problem of relay selection, channel access and power allocation in CR systems where the system state changes according to a Markov chain. Due to the imperfect sensing, the channel state information is uncertain in CR systems and a learning stochastic optimization approach has to be developed to derive the most likely state and make a stochastic control. Although some other methods, like dual decomposition [24] and interior point [25], can be used to solve the optimization problem, these methods cannot deal with the uncertainty. In this paper, we formulate the stochastic optimization problem in CR system as a POMDP problem which has been widely used to deal with stochastic control issues with uncertainty, such as artificial intelligence research [26], dynamic sensor scheduling [27], and transmission control in wireless networks [7]. Since the system state cannot be fully observed owing to the imperfect observation, an information state that encapsulates the decision and observation history is thus introduced to infer the most likely



policy that determines the action to be taken. In what follows, we will formulate the CR system as a POMDP problem and investigate its solution.

#### A. POMDP Formulation

A POMDP is defined by a hextuple,  $\langle \mathcal{A}, \mathcal{S}, \mathcal{P}, \Theta, \mathcal{B}, \mathcal{R} \rangle$ , where  $\mathcal{A}$ ,  $\mathcal{S}$ ,  $\mathcal{P}$ ,  $\Theta$ ,  $\mathcal{B}$ , and  $\mathcal{R}$  denote the action to be taken, the system state, the transition probability for each action in each state, the finite set of observations, the observation model, and the reward structure, respectively.

1) *Action Space*: The action space is a combined action space, in which the channel sensing decision, sensor operating point selection, access decision are included. The sequence of operations [19] in each slot is: Sensing Decision, Sensing Observation, Access Decision, Power Allocation, and Reward. The composite action at slot is defined as  $a = [a_s, a_{\epsilon, \delta}, a_a, a_p]$  ( $a \in \mathcal{A}$ ), where  $a_s$ ,  $a_{\epsilon, \delta}$ ,  $a_a$ , and  $a_p$  denote the sensing action, sensor operating point selection, access decision, and power allocation, respectively. Note that, the symbols  $a_{(*)}(k)$  present the corresponding actions at slot  $k$  throughout the rest of this paper.

**Sensing Decision:** At the beginning of slot  $k$ , based on its current information state, the system first determines the operating mode in this slot: sleeping or sensing. If the sleeping mode is selected, no further decision needs to be made in this slot. Otherwise, the system chooses a potential relay  $R_n$  to sense its surrounding channel  $l$ . In this paper, we consider that each secondary user is equipped with a single Neyman-Pearson energy detector [28], and can only sense one channel at a time. Let “0” represent the sleeping state. The sensing action,  $a_s(k)$ , is defined as

$$a_s(k) \in \{0(no\_sensing), ch_1^{R_1}, \dots, ch_L^{R_1}, ch_1^{R_2}, \dots, ch_L^{R_N}\}. \quad (4)$$

**Sensor Operating Point Selection:** The receiver operating characteristic (ROC) curve is widely used to represent the performance of the sensor, and displays the tradeoff between the probability of false alarm and the probability of miss detection [19]. In this paper, let  $\delta$  and  $\epsilon$  denote the probability of miss detection (the probability that fails to detect the presence of

primary users) and the probability of false alarm (the probability that the sensed channel is considered to be in use whereas it is actually off), respectively. For a communication system with the constraint of collision probability, the spectrum sensor operating point is set such that  $\delta = \zeta$ , where  $\zeta$  is the required collision probability. Therefore, the sensor operating action,  $a_{\epsilon, \delta}$ , is then determined.

**Access Decision:** After observing  $\theta(k)$  from the chosen channel, the user determines whether or not to access it. The access decision,  $a_a(k)$ , is defined as

$$a_a(k) \in \{0(no\_access), 1(access)\}. \quad (5)$$

In CR systems, secondary users cannot interfere with primary users. Thus, the opportunity of accessing a channel has to be given up if this channel is inferred to be in use by a primary user.

**Power Allocation Decision:** When a channel is selected, the corresponding power allocation decision will be made based on the goal of achieving the maximum received reward. Let  $a_p(k)$  denote the power allocation decision as

$$a_p(k) \in [0, P^{mtx}]. \quad (6)$$

2) *State Space and Transition Probability:* In CR systems, for a potential relay node, its state is characterized by the channel state information and the residual energy state information. We consider that each potential relay node can sense only one of  $L$  channels. In this system with  $N$  potential relay nodes, the total number of channels is  $M = L \cdot N$  and the index of channel  $l$  for relay  $R_n$  is  $(n-1) \cdot L + l$ . The system state is presented by all channel states and thus is denoted by  $S^k = [s_1^k, s_2^k, \dots, s_M^k]$  in slot  $k$ , where  $s_m^k$  is the state of channel  $m$ , and the probability of the state transition is denoted by

$$p_m(i, j) = \Pr\{s_m^{k+1} = j | s_m^k = i\}, \forall i, j \in \mathcal{C}. \quad (7)$$

The transition matrix of the system state is given by an  $(|S| + 1)^M \times (|S| + 1)^M$  matrix  $T$ . Therefore, the system state evolves according to the following transition probability matrix

$$P = [p_1(i, j), p_2(i, j), \dots, p_M(i, j)]_{S \times S}. \quad (8)$$

3) *Observation Space*: It is unable to obtain the full knowledge of channel state in each slot due to the false alarm and miss detection. Suppose that the system has decided to allow a potential relay to sense its corresponding channel in a slot. Then, the secondary user observes the channel state. The observation available to the secondary user is the sensed channel outcome. Let  $\theta \in \Theta$  represent the observed state, where  $\Theta = \{h_0, h_1, \dots, h_i, \dots, h_S\}$  ( $i \in \{0, 1, \dots, S\}$ ).

Let  $b_{j,\theta}^a = \Pr\{\theta|j, a\}$  denote the probability that the observed state is  $\theta$  given that the channel is in state  $j$  and composite action  $a$  in the last slot was taken. Hence, we can obtain the conditional probability of observing  $\theta$ ,

$$b_{j,\theta}^a = \begin{cases} 1 - \delta, & \text{if } \theta = h_0, j = 0, \\ \delta, & \text{if } \theta \neq h_0, j = 0, \\ \epsilon, & \text{if } \theta = h_0, j \neq 0, \\ (1 - \epsilon), & \text{otherwise,} \end{cases} \quad (9)$$

where  $v(\theta)$  is a function to determine the channel state when the channel gain is obtained, and  $v(\theta) = i$ ,  $0 \leq i \leq S$  given  $\theta = h_i$ .

4) *Information State*: Information state is an important concept in POMDP systems. Although the system state cannot be directly known, it can be inferred from its decision and observation history encapsulated by the information state. The information state is a probability distribution over the state space. Let  $\pi_k = \{\pi_k^1, \pi_k^2, \dots, \pi_k^{(|S|+1)^M}\}$  denote the information space, and  $\pi_k^i \in [0, 1]$  represent the conditional probability (given decision and observation history) that the system is in state  $i \in S$  at the beginning of slot  $k$  prior to state transition.

At the end of each slot, the information state is updated using Bayes' rule [29]:

$$\pi_{k+1}^i = \frac{\sum_{i'} \pi_k^{i'} p_{i',i} b_{i\theta}^a}{\sum_{i',i} \pi_k^{i'} p_{i',i} b_{i\theta}^a}. \quad (10)$$

5) *Reward and Objective:* In CR systems, the achievable system capacity and the energy consumption are the important parameters for performance evaluation. Therefore, the system reward of the proposed scheme consists of two parts: 1) the instantaneous received reward contributed by the achievable rate, and 2) the immediate cost incurred by the energy consumption. The reward function is originally defined as

$$R(R_n, l, P_R^{tx}) = \begin{cases} 0, & \text{if there is no sensing,} \\ \alpha \cdot C_{AF}(R_n, l, P_R^{tx}) - \eta \cdot C_{EN}(R_n, l, P_R^{tx}), & \text{otherwise,} \end{cases} \quad (11)$$

where  $R(R_n, l, P_R^{tx})$  is the received reward when channel  $l$  at relay  $R_n$  is selected to transmit information at the expense of power consumption  $P_R^{tx}(k)$  in slot  $k$ ;  $C_{AF}(R_n, l, P_R^{tx})$  and  $C_{EN}(R_n, l, P_R^{tx})$  are the immediate reward and cost in slot  $k$ , respectively;  $\alpha$  and  $\eta$  ( $0 \leq \alpha, \eta \leq 1$ , and  $\alpha + \eta = 1$ ) are coefficient. They are used to adjust the effect of immediate reward and cost on the total reward, and ultimately affect the joint strategy of relay selection, channel access, and power allocation. As pointed out in [18], the immediate reward and cost can be well balanced if they are at the same scale.

$C_{AF}(R_n, l, P_R^{tx})$  is the system capacity calculated by Eq. (1) while  $C_{AF}(R_n, l, P_R^{tx})$  is related to the energy consumption and is defined as

$$C_{EN}(R_n, l, P_R^{tx}) = f(P_R^{tx}) = e^{\frac{P_R^{cs} \cdot t_{cs} + P_R^{tx} \cdot t_{tx}}{P^{mtx} \cdot T}} - 1. \quad (12)$$

$C_{EN}(R_n, l, P_R^{tx})$  is the cost function to evaluate the effect of energy consumption for information transmission on the immediate cost. The energy consumption for sensing and information transmission are jointly considered in this function, and is normalized by the upper bounded energy consumption,  $P^{mtx} \cdot T$ , so as to make the value of reward and cost in the same scale.

The objective of the proposed scheme is to maximize the expected total reward for  $K$  slots under the constraint that the collision probability is bounded below  $\zeta$  by developing a combined policy in CR systems. Hence, the optimal policy can be derived from the following equation

$$\mu^* = \arg \max_{\{\mu_s, \mu_e, \delta, \mu_a, \mu_p\}} E \left[ \sum_{k=1}^K R(R_n, l, P_R^{tx}) \right],$$

subject to

$$P_c(k) = \Pr\{a_a(k) = 1 | s_m^k = 0\} < \zeta, \forall k \in \{1, 2, \dots, K\}, \quad (13)$$

where  $m = (n - 1) \cdot L + l$ ;  $\mu_s$  is a channel sensing policy which specifies the sensing decision  $a_s$ ;  $\mu_{\epsilon, \delta}$  is a sensor operating policy that specifies a spectrum sensor design  $(\epsilon, \delta) \in \mathbb{A}_{\epsilon, \delta}$  based on the system tolerable probability of collision  $\zeta$ , where  $\mathbb{A}_{\epsilon, \delta}$  are the valid point on the ROC curve;  $\mu_a$  is an access policy that specifies the access decision  $a_a$ ;  $\mu_p$  is a power allocation policy that specifies the power allocation decision  $a_p$ ;  $E_{\{\mu_s, \mu_{\epsilon, \delta}, \mu_a, \mu_p\}}$  indicates the expectation given that policies  $\mu_s, \mu_{\epsilon, \delta}, \mu_a, \mu_p$  are employed; The symbols  $\mu_{(\cdot)}^*$  mean the corresponding optimal policies derived from Eq. (13).

### B. The Solution to the POMDP Problem

This subsection will develop an optimal policy for the joint relay selection, channel access, and power allocation problem in CR systems based on the POMDP formulation, which determines the relay to be selected, the channel to be accessed, and the corresponding transmission power. To solve the POMDP problem, the dynamic programming method and dimension reduction strategy are used. The dynamic programming is firstly proposed to solve the POMDP problem where the value function is introduced to derive the optimal policy, but its computational complexity exponentially increases with the number of states in a state space. Therefore, a dimension reduction strategy is then proposed to derive a suboptimal policy by exploiting the marginal distribution, and consequently to low the computation complexity of solving the POMDP problem.

*1) Dynamic Programming to the POMDP formulation:* The design objective is to determine which composite action is taken so that the expected total reward obtained during  $K$  slots is maximized. In order to effectively calculate Eq. (13), the dynamic programming is applied to compute the optimal policy. Referred to as the value function,  $J_k(\pi_k)$  denotes the maximum expected remaining reward that can be accrued starting from slot  $k$  when the current belief

vector is  $\pi_k$ . The value function includes two parts: 1) the immediate reward obtained in slot  $k$  which is given by  $R(s_m^k)$  when the network is in state  $j$  and the user takes composite action  $a_k$  and observes  $\theta_k$ ; 2) the maximum expected remaining reward  $J_{k+1}(\pi_{k+1})$  starting from slot  $k+1$  given a belief vector  $\pi_{k+1} = U(\pi_k|a_k, \theta_k)$  that represents the updated knowledge of the system state after incorporating the action  $a_k$  and the observation  $\theta_k$  at time slot  $k$ . Averaging over all possible system states and observations, we have the following Bellman's equation.

$$J_k(\pi_k) = \max_{a \in A} \sum_{i \in \mathcal{S}} \sum_{i' \in \mathcal{S}} \pi_j^k p_{i',i} \sum_{j=1}^{j=\mathcal{S}} b_{j,\theta_k}^{a_k} [R(R_n, l, P_R^{tx}) + J_{k+1}(U(\pi_k|a_k, \theta_k))], 1 \leq k \leq K. \quad (14)$$

It can be seen from Eq. (14) that the selected policy in a slot affects the total reward in two ways: 1) how to obtain an immediate reward  $R(R_n, l, P_R^{tx})$ ; and 2) how to transform the belief vector to  $U(\pi_k|a_k, \theta_k)$  which determines the future reward  $J_{k+1}(U(\pi_k|a_k, \theta_k))$ .

2) *Dimension Reduction Strategy*: To obtain an optimal policy taken at each time slot, we have to compute the value function that is hinged on the optimal policy. However, since the information state space grows exponentially with the number of the total mapped channels, the computation of optimal policy is still difficult to be derived even if it can be computed off-line. Fortunately, an alternative information state is given by the marginal distribution [30] for this special Markov process that the system state evolves independently of the previous actions. The dimension of the information state space can be reduced to grow linearly with  $M$ , but the policy is just suboptimal.

$\Omega_k = [\omega_1(k), \dots, \omega_M(k)]$  is an information state vector for the optimal policy under  $M$  independent channels, where  $\omega_l(k)$  is a vector that is the probability over channel state space  $\mathcal{S}$  (based on the entire decision and observation history) at the beginning of slot  $k$ , and  $\omega_m(k) = [\omega_m^1(k), \omega_m^2(k), \dots, \omega_m^S(k)]$ .

The belief is updated in a similar way as done in Eq. (10). The updating equation follows

$\Omega_{k+1} = U(\Omega_k)$ , where

$$\begin{aligned} \omega_m^i(k) = & \omega_m^1(k-1) \Pr\{s_m^k = i | s_m^{k-1} = 1\} + \dots + \omega_m^j(k-1) \Pr\{s_m^k = i | s_m^{k-1} = j\} + \dots \\ & + \omega_m^S(k-1) \Pr\{s_m^k = i | s_m^{k-1} = S\}. \end{aligned} \quad (15)$$

Once the probability that channel index  $m$  (i.e., it refers to channel  $l$  of relay  $R_n$ ) is in state  $s_m^k$  at slot  $k$  is acquired, the expected reward  $R_m^{s^k}(k)$  at state  $i$  is obtained and the equation is as follow

$$R_m^{s^k}(k) = \omega_m^k \cdot R(R_n, l, P_R^{tx}). \quad (16)$$

The action in slot  $k$  is to maximize the expected immediate reward and the index  $a^*(k)$  of the selected channel is given by

$$a^*(k) = \arg \max_{m=1,2,\dots,M} \max_{s^k \in \mathcal{C}} R_m^{s^k}(k). \quad (17)$$

At the end of time slot  $k$ , the belief vector  $\Omega$  is updated based on the action  $a^*(k)$  and the observation  $\Theta_{a^*}(k)$  as follows.

$$\Omega_{k+1} = [\omega_1(k+1), \dots, \omega_M(k+1)] = U(\Omega_k | a^*(k), \Theta_{a^*}(k)), \quad (18)$$

where

$$\omega_m(k+1) = \begin{cases} 0, & \text{if } a^*(k) = m, \Theta_{a^*}(k) = h_0, \\ \Pr\{s_m^{k+1} = i\}, & \text{if } a^*(k) = m, \Theta_{a^*}(k) = h_i, \\ \omega_m^1(k) \Pr\{s_m^{k+1} = i | s_m^k = 1\} + \omega_m^2(k) \Pr\{s_m^{k+1} = i | s_m^k = 2\} + \dots \\ \quad + \omega_m^S(k) \Pr\{s_m^{k+1} = i | s_m^k = S\}, & \text{if } a^*(k) \neq m. \end{cases} \quad (19)$$

Let  $W_k(\Omega)$  denote the expected remaining reward starting from slot  $k$ . A recursive equation for  $W_k(\Omega)$  can be obtained as follows.

$$W_k(\Omega) = R_{a^*}(k) + \sum_{\theta=h_0}^{\theta=h_S} \Pr[\Theta_{a^*} = \theta | \Omega, a^*] W_{k+1}(U(\Omega | a^*, \theta)), \quad (20)$$

where  $R_{a^*}(k) = \omega_{a^*}^k \cdot R(s_{a^*}^k)$ .

3) *The Solution to the POMDP after the Dimension Reduction:* This value function with finite action space can be solved using linear programming techniques [31]. Smallwood and Sondik [29] showed that value function  $W_k(\Omega_k)$  is piecewise linear and convex. The domain of  $W_k(\Omega_k)$  can be partitioned into a finite number of convex regions  $C_1(k), \dots, C_N(k)$ . Associated with each region  $C_i(k)$  is a vector  $\Upsilon_i(k)$  such that value function  $W_k(\Omega_k)$  in this region is given by the inner product of  $\Omega_i^k$  ( $\Omega_i^k \in C_i(k)$ ) and  $\Upsilon_i(k)$ . Therefore, after using this structure, the following equation can be obtained.

$$W_k(\Omega_k) = R_{a^*}(k) + \sum_{\theta=h_1}^{\theta=h_S} \Pr[\Theta_{a^*} = \theta | \Omega_k, a^*] * \langle \Omega_{k+1}, \Upsilon_{i_{\Omega_{k+1}}}(k+1) \rangle, 1 \leq k \leq K-1, \quad (21)$$

where  $\langle \cdot, \cdot \rangle$  denotes inner product and  $i_{\Omega_{k+1}}$  is the index of the region containing the updated belief vector  $\Omega_{k+1} = U(\Omega_k | a_k, \theta_k)$ . Thus, if the convex regions  $\{C_i(k+1)\}$  and the associated  $\Upsilon$ -vector  $\Upsilon_i(k+1)$  have been calculated for slot  $k+1$ , we can obtain the optimal actions and the corresponding  $c$ -vectors for slot  $t$  from Eq. (21). The optimal policy is thus given by the convex regions and the associated  $c$ -vectors and optimal actions for  $k = 1, \dots, K$ .

4) *Finding the Upper Bound:* The optimality of the proposed scheme is further analyzed in this part. By considering a genie aided system, the upper bound of the scheme is established to evaluate the performance loss. In the genie aided system, each relay can sense, access, and obtain observations from all  $L$  channels. The observation,  $h_i$ , is perfect and the channel gain is accurate. Moreover, the source can get all information from the relay candidates as well. Therefore, the policy of joint relay selection, channel access, and power allocation is perfect in each slot. The obtained reward,  $E_{genie}^n$ , is

$$E_{genie}^n = R(R_n, l, P_R^{tx}), \quad (22)$$

and the maximum reward,  $E_{genie}$ , is

$$E_{genie} = \max \{E_{genie}^1, E_{genie}^2, \dots, E_{genie}^N\}. \quad (23)$$



$E_{genie}$  is the reward obtained in the genie aided system, and is also the upper bound of the proposed scheme in this paper.

It is obvious that the upper bound obtained in the genie aided system is loose. Thus, a tighter upper bound is further established. In CR systems, the optimal system performance is bounded by the fact that a relay that can achieve the maximum reward is unable to be always selected to relay information. This is because only a relay is selected to sense channel and forward information in the proposed scheme. Therefore, the upper bound is that all relay candidates can sense a channel and derive a optimal policy to determine its channel access and power allocation and the one that can achieve the maximum expected reward can be selected as a relay. For a relay candidate  $R_n$ , the optimal policy of joint channel access and power allocation,  $(a_{n,a}^*, a_{n,p}^*)$  is

$$(a_{n,a}^*, a_{n,p}^*) = \arg \max E_{\{\mu_s, \mu_{\epsilon}, \mu_a, \mu_p\}} \left[ \sum_{k=1}^K R(R_n, l, P_R^{tx}) \right], \quad (24)$$

and the maximum expected reward for candidate  $R_n$ ,  $E_{max}^n$ , is

$$E_{max}^n = \max E_{\{\mu_s, \mu_{\epsilon}, \mu_a, \mu_p\}} \left[ \sum_{k=1}^K R(R_n, l, P_R^{tx}) \right].$$

The source will sort these expected reward from all candidate relays and choose the best one, that is,

$$E_{max} = \max \{E_{max}^1, E_{max}^2, \dots, E_{max}^N\}. \quad (25)$$

The optimal policy for each relay candidate is derived by the POMDP process and the expected reward is obtained by using value function as done by Eq. (14).

### C. The Centralized Decision Framework

In CR systems, the channel condition cannot be perfectly sensed due to the false alarm and the miss detection. As a consequence, the system capacity may have a sharp decline if the relay and channel are scheduled as well as the transmission power is allocated according to the sensed output. A learning based scheme is proposed to address this problem and the most likely

channel state is deduced. In CR systems, source, destination and relay nodes shatter over the region. In this scenario, it is necessary to consider carefully how to make a decision. In this paper, a centralized decision framework is propose, where the decision is made at the source.

In this framework, each candidate relay senses the designated spectrum, and then sends the outcome back to the source node. The detailed process is described as follows.

1) The process of forming a potential relay set:

- a) A source node broadcasts a pilot message, INIT\_BEACON, to determine its neighbors.
- b) Each node received the message will reply with POS\_RESP to indicate that it is a neighbor.
- c) The source node send to a group of neighbors a multicast message, REL\_REQ, asking whether they are able to relay information.
- d) The neighbor that is willing to be a candidate relay sends a message, REL\_JOIN\_REQ, to the source for joining the relay set.
- e) Finally, the source will return an acknowledgement message. Thus, the potential relay set is formed.

2) The process of making a decision:

- a) When a source node transmits information to its intended destination, the decision system also firstly infer the sensing decision that determine whether a candidate relay in the potential relay set needs to sense the channel. If a candidate relay is selected to sense the channel, the sensing request message, REL\_SENSE\_REQ, will be sent.
- b) After sensing the channel for a while (i.e.,  $t_{cs}$ ), the selected relay then sends the sensed outcome,  $\gamma$  (i.e., the observation  $h_i$ ), to the decision system.
- c) Then, the channel access and its corresponding power allocation decision are derived and the decision message, INIT\_TRANS\_REQ, is sent to all relays through using

multicast transmission.

- d) At the end of the slot, the accurate channel gain and residual energy are updated as well as the decision history is recorded.

#### IV. SIMULATION RESULTS AND PERFORMANCE ANALYSIS

This section presents simulation results to evaluate the performance of the proposed scheme. The simulation is executed in the Rayleigh block-fading wireless communication environment. The network scenario is depicted in Fig. 1, where **S** transmits information to the corresponding **D** with the help of **R** selected from the relay set  $R$ . The AF protocol is applied to support CC. The simulation area is a  $10 \text{ kilometer} \times 10 \text{ kilometer}$  square. The Manhattan mobility model [32] is adopted in the simulation and relays move according to this model at a low speed. The battery capacity for each relay is 1000 mAh with the output voltage 1 Volt. The maximum transmission power for the source and relay is set as  $P^{mtx} = 150mw$ . Within a time slot, 10% of the time is used for channel sensing and the other time is for data transmission. Moreover, only one potential relay and its corresponding channel are selected to be accessed. The coefficient  $\alpha$  and  $\eta$  are set to be  $\alpha = \eta = 0.5$  and the discount factor is set to be  $\beta = 0.8$ .

For a channel, the probability of occupancy by primary users can be obtained after a long time statistics. For a channel  $l$  ( $l \in \{1, 2, \dots, L\}$ ), the channel state transitioning matrix is constructed by the probability that the channel stays at busy state  $\Pr\{s_l^{k+1} = z | s_l^k = z\} = 0.05$ , the probability of channel transitioning into busy state  $\Pr\{s_l^{k+1} = z | s_l^k = v\} = 0.15$ , and the probability of dwelling into the same available state  $\Pr\{s_l^{k+1} = v | s_l^k = v\} = 0.9$ . The above parameters are widely used in the existing related studies [19], [15]. The channel sensing false alarm probability is  $\epsilon = 0.1$ , and the channel sensing miss detection probability is  $\delta = 0.1$ . In order to investigate and compare the performance of the proposed scheme thoroughly, three cases are taken into account: 1) the proposed scheme that exploits the learning approach and considers the energy consumption, 2) the existing scheme that no learning approach is considered, and 3)

the existing scheme that no energy is considered. The performance evaluation and comparison are conducted in terms of the average received reward, average system capacity, and average network lifetime. It is noteworthy that the energy consumption can be reflected by the average network lifetime.

#### *A. The Performance Improvement by the Proposed Scheme*

Figs. 2-4 depict the average received reward, system capacity, and the network lifetime for different schemes, respectively. The used parameters are set the same as described above, and the channel state is changed in this simulation experiment.

Fig. 2 illustrates the reward improvement that is relative to the system capacity and energy consumption by the proposed scheme compared to the two existing schemes. It can be seen that the proposed scheme has the highest reward, while the received reward of the existing scheme without energy consideration is higher than that of the existing scheme without learning consideration. The results highlight the importance of jointly considering the relay selection, channel access, and power allocation as well as exploiting the learning based scheme in CR systems with imperfect sensing.

Fig. 3 compares the system capacity achieved by the proposed scheme with those by the two existing schemes. As shown in this figure, the system capacity achieved by the proposed scheme is slightly lower than that of the existing scheme without energy consideration but higher than that of the learning based existing scheme. This is because the joint relay selection, channel access, and power allocation is under the goal of maximizing the system capacity in the existing scheme without energy consideration. Besides, due to no learning in CR systems, the decision of the relay selection, channel access, and power allocation is made accurately, which can incur the degradation of system capacity. From this figure, it can be concluded that the learning based scheme has to be applied in CR systems with imperfect sensing.

Fig. 4 depicts the network lifetime with different schemes. The network lifetime of the

proposed scheme is the highest among all these schemes because the energy consumption is considered in the proposed scheme. Moreover, the network lifetime of the existing scheme without energy consideration is higher than that of the existing scheme without learning consideration. Furthermore, more energy is wasted because more collisions between primary users and secondary users occur in the existing scheme without the consideration of learning. The figure reveals that the energy consumption should be considered and the balance between the system capacity and network lifetime should be achieved as well.

These figures show that the received reward increases more than 50%, the system capacity is reduced less than 6%, and the network lifetime is increased more than 35% owing to the joint relay selection, channel access, and energy consumption scheme proposed in this paper. Meanwhile, the system performance can be increased through making use of the learning based scheme in CR systems with imperfect sensing as well. Besides, the network lifetime can be improved while the system capacity is degraded through reducing the energy consumption, which inspires the consideration of the tradeoff between the system capacity and power allocation.

### *B. Effects of Channel Sensing Capacity*

This subsection investigates how the channel sensing capacity affects the reward and system capacity. Figs. 5 and 6 investigate the effects on the received reward and system capacity, respectively. The channel sensing miss detection probability is  $\delta = 0.1$ . The probability of the false alarm changes in this simulation experiment. Note that, as shown by the receiver operating characteristic (ROC) curve, there is a relationship between the probability of false alarm and miss detection. Hence, it is sufficient to consider the effect of the false alarm here.

Figs. 5 and 6 reveal that the received reward and system capacity can be significantly affected by the channel sensing capacity in CR systems. As shown in Fig. 5, the received reward and system capacity decrease as the probability of false alarm increases. Furthermore, the proposed scheme and the scheme without energy consideration are all better than the scheme without

learning consideration. The reason is that the increase in the false alarm probability causes the increase of abandoning the high-quality relay and corresponding channel, and consequently reducing the received reward and system capacity. This trend indicates that the channel sensing capacity has significant impact on the system performance.

### *C. Effects of the Return of Primary Users*

Figs. 7-9 illustrate the effects of the return of primary users on the received reward, system capacity, and network lifetime under different schemes, respectively. In order to evaluate how the activity of primary users affects the system performance, the probability of the channel transitioning from available state into busy state is varied.

Figs. 7 and 8 show that the received reward and system capacity degrade with the increase in the probability of the primary user's return, whilst Fig. 9 reveals the increase in the network lifetime. The reason is that when the primary user occupies a channel in a slot, the secondary user has to give up the channel access. As a consequence, there is no reward and system capacity to be received in this slot. It is obvious that the higher the probability of primary users' return, the few opportunities for the relay's transmission. On the contrary, the energy consumption reduces for no transmission, which means that the network lifetime will be prolonged. It can also be seen from Figs. 7-9 that the received reward of the proposed scheme is higher than that of the existing schemes. However, the system capacity is lower than that of the existing scheme without energy consideration, and higher than that of the existing scheme without learning.

## V. CONCLUSIONS AND FUTURE WORK

In this paper, an energy-aware relay selection scheme has been proposed to achieve the goal of balancing the system capacity and the energy consumption as well as maximizing the system reward. This scheme jointly takes the relay selection, channel access, and power allocation into account in cognitive radio (CR) systems with imperfect sensing. To achieve this goal, the problem of relay selection, channel access, and power allocation has been investigated and the CR systems

with imperfect sensing has been formulated as a partially observable Markov decision process (POMDP). The optimal policy for relay selection, channel access, and power allocation is derived by a dynamic programming approach and the computation complexity is reduced by applying the dimension reduction strategy. The upper bound is derived to evaluate the performance loss of the proposed scheme. Extensive simulation experiments and results have been presented to demonstrate the significant performance improvement compared to the existing scheme. Future work is in progress to consider other important issues, such as multimedia transmission in CR systems, in the proposed framework.

## REFERENCES

- [1] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "Next generation/dynamic spectrum access/cognitive radio wireless networks: A survey," *Computer Networks*, vol. 50, no. 13, pp. 2127–2159, Sept. 2006.
- [2] E. Axell, G. Leus, E. G. Larsson, and H. V. Poor, "Spectrum sensing for cognitive radio: State-of-the-art and recent advances," *IEEE Signal Processing Mag.*, vol. 29, no. 3, pp. 101–116, May 2012.
- [3] H. Kim and K. G. Shin, "Efficient discovery of spectrum opportunities with MAC-layer sensing in cognitive radio networks," *IEEE Trans. Mobile Comput.*, vol. 7, no. 5, pp. 533–545, May 2008.
- [4] S. Haykin, "Cognitive radio: Brain-empowered wireless communications," *IEEE J. Select. Areas Commun.*, vol. 23, no. 2, pp. 201–220, Feb. 2005.
- [5] G. Bansal, J. Hossain, and V. K. Bhargava, "Optimal and suboptimal power allocation schemes for OFDM-based cognitive radio systems," *IEEE Trans. Wireless Commun.*, vol. 7, no. 11, pp. 4710–4718, Nov. 2008.
- [6] A. D. Domenico, E. C. Strinati, and M. D. Benedetto, "A survey on MAC strategies for cognitive radio networks," *IEEE Commun. Surveys & Tutorials*, vol. 14, no. 1, pp. 21–44, 1st Quarter. 2012.
- [7] C. Luo, F. R. Yu, H. Ji, and V. C. M. Leung, "Cross-layer design for TCP performance improvement in cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 59, no. 5, pp. 2485–2495, Jun. 2010.
- [8] L. Fu, Y. J. A. Zhang, and J. Huang, "Energy efficient transmissions in mimo cognitive radio networks," *IEEE J. Select. Areas Commun.*, vol. 31, no. 11, pp. 2420–2431, Nov. 2013.
- [9] N. Laneman, D. Tse, and G. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inform. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.
- [10] Y. Hong, W. Huang, F. Chiu, and J. Kuo, "Cooperative communications in resource-constrained wireless networks," *IEEE Signal Processing Mag.*, vol. 24, no. 3, pp. 47–57, May 2007.
- [11] O. Simeone, Y. Bar-Ness, and U. Spagnolini, "Stable throughput of cognitive radios with and without relaying capability," *IEEE Trans. Commun.*, vol. 55, no. 12, pp. 2351–2360, Dec. 2007.
- [12] P. Gong, J. Park, J. Yoo, B. Yu, and D. Kim, "Throughput maximization with multiuser non-selfish cognitive relaying in cognitive radio networks," in *Proc. IEEE ISWPC'09*, Melbourne, Australia, Feb. 11–13 2009, pp. 1–5.
- [13] J. Jia, J. Zhang, and Q. Zhang, "Cooperative relay for cognitive radio networks," in *Proc. IEEE INFOCOM'09*, Rio de Janeiro, Brazil, Apr. 19–25 2009, pp. 2304–2312.
- [14] G. Zhao, C. Yang, G. Li, D. Li, and A. Soong, "Power and channel allocation for cooperative relay in cognitive radio networks," *IEEE J. Select. Topics Signal Processing*, vol. 5, no. 1, pp. 151–159, Feb. 2011.
- [15] L. Li, X. Zhou, H. Xu, G. Y. Li, D. Wang, and A. Soong, "Simplified relay selection and power allocation in cooperative cognitive radio systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 1, pp. 33–36, Jan. 2011.
- [16] J. Liu, W. Chen, Z. Cao, and Y. J. A. Zhang, "Cooperative beamforming for cognitive radio networks: A cross-layer design," *IEEE Trans. Commun.*, vol. 60, no. 5, pp. 1420–1431, May 2012.
- [17] P. Li, S. Guo, W. Zhuang, and B. Ye, "Capacity maximization in cooperative CRNs: Joint relay assignment and channel allocation," in *Proc. IEEE ICC'12*, Ottawa, ON Canada, Jun. 10–15 2012, pp. 5097–5101.

- [18] C. Luo, G. Min, F. R. Yu, M. Chen, L. T. Yang, and V. C. M. Leung, "Energy-efficient distributed relay and power control in cognitive radio cooperative communications," *IEEE J. Select. Areas Commun.*, vol. 31, no. 11, pp. 2442–2452, Nov. 2013.
- [19] Y. Chen, Q. Zhao, and A. Swami, "Joint design and separation principle for opportunistic spectrum access in the presence of sensing errors," *IEEE Trans. Inform. Theory*, vol. 54, no. 5, pp. 2053–2071, May 2008.
- [20] Y. Chen, S. Zhang, S. Xu, and G. Y. Li, "Fundamental trade-offs on green wireless networks," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 30–37, Jun. 2011.
- [21] T. Han and N. Ansari, "On greening cellular networks via multicell cooperation," *IEEE Wireless Commun. Mag.*, vol. 20, no. 1, pp. 82–89, Feb. 2013.
- [22] Q. Zhang and S. Kassam, "Finite-state Markov model for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, no. 11, pp. 1688–1692, Nov. 1999.
- [23] C.-D. Iskander and P. T. Mathiopoulos, "Fast simulation of diversity Nakagami fading channels using finite-state Markov models," *IEEE Trans. Broadcast.*, vol. 49, no. 3, pp. 269–277, Sept. 2003.
- [24] D. Hu and S. Mao, "On medium grain scalable video streaming over femtocell cognitive radio networks," *IEEE J. Select. Areas Commun.*, vol. 30, no. 3, pp. 641–651, Apr. 2012.
- [25] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U. K.: Cambridge University Press, 2004.
- [26] A. Cassandra, M. Littman, and N. Zhang, "Incremental pruning: A simple fast exact method for partially observed markov decision processes," in *Proc. 13th Annual Conference on Uncertainty in Artificial Intelligence*, San Francisco, CA, USA, Aug. 1-3 1997, pp. 54–61.
- [27] V. Krishnamurthy and D. V. Djonin, "Structured threshold policies for dynamic sensor scheduling— a partially observed markov decision process approach," *IEEE Trans. Signal Processing*, vol. 55, no. 10, pp. 4938–4957, Oct. 2007.
- [28] A. Patel and B. Kosko, "Optimal noise benefits in Neyman-Pearson and inequality-constrained statistical signal decision," *IEEE Trans. Signal Processing*, vol. 57, no. 5, pp. 1655–1669, May 2009.
- [29] R. D. Smallwood and E. J. Sondik, "Optimal control of partially observable Markov processes over a finite horizon," *Operations Research*, vol. 21, pp. 1071–1088, 1973.
- [30] Q. Zhao, L. Tong, A. Swami, and Y. Chen, "Decentralized cognitive MAC for opportunistic spectrum access in Ad Hoc networks: A POMDP framework," *IEEE J. Select. Areas Commun.*, vol. 25, no. 3, pp. 589–600, Apr. 2007.
- [31] W. Lovejoy, "A survey of algorithmic methods for partially observed Markov decision processes," *Ann. of Oper. Res.*, vol. 28, no. 1, pp. 47–66, 1991.
- [32] Y. Lu, H. Lin, Y. Gu, and A. Helmy, "Towards mobility-rich analysis in Ad Hoc networks: Using contraction, expansion and hybrid models," in *Proc. IEEE ICC'04*, Paris, France, Jun. 20-24 2004, pp. 4346–4351.



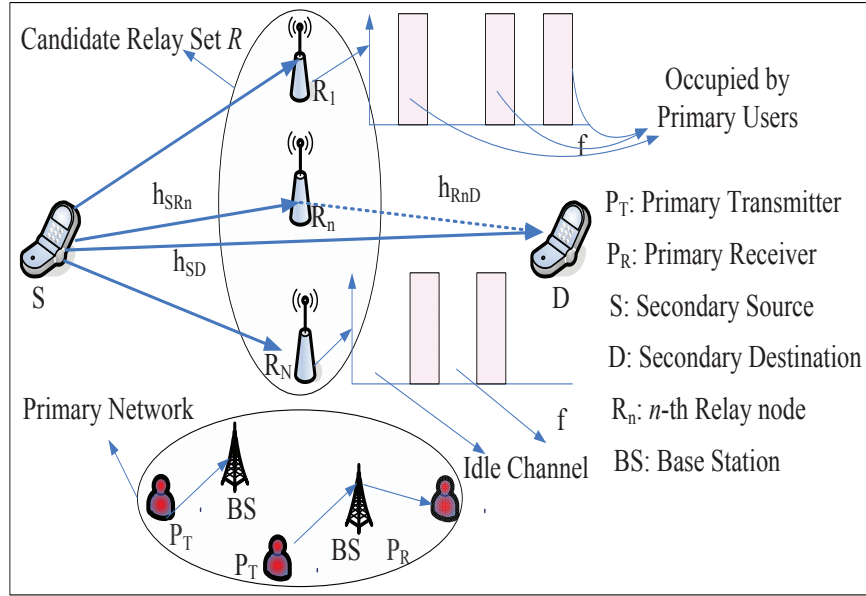


Fig. 1. The CR system model where the communication between source and destination is aided by a relay in the potential relay set.

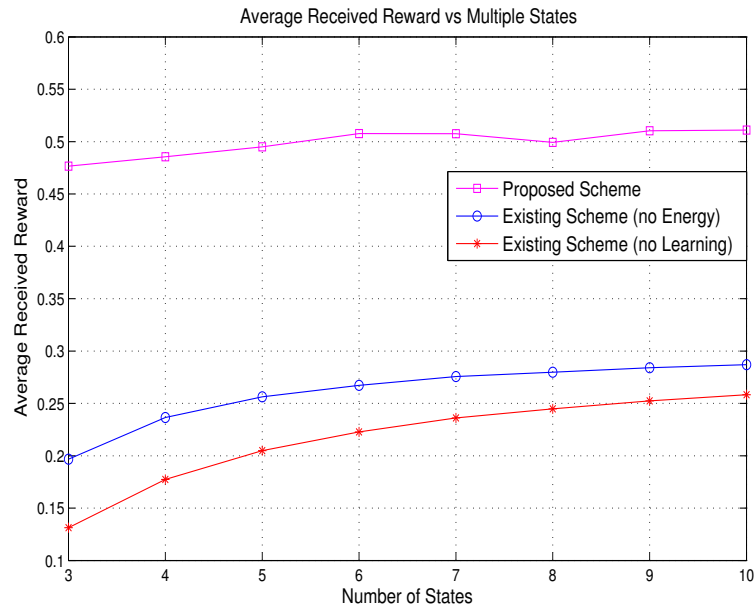


Fig. 2. Average received reward versus the number of states.

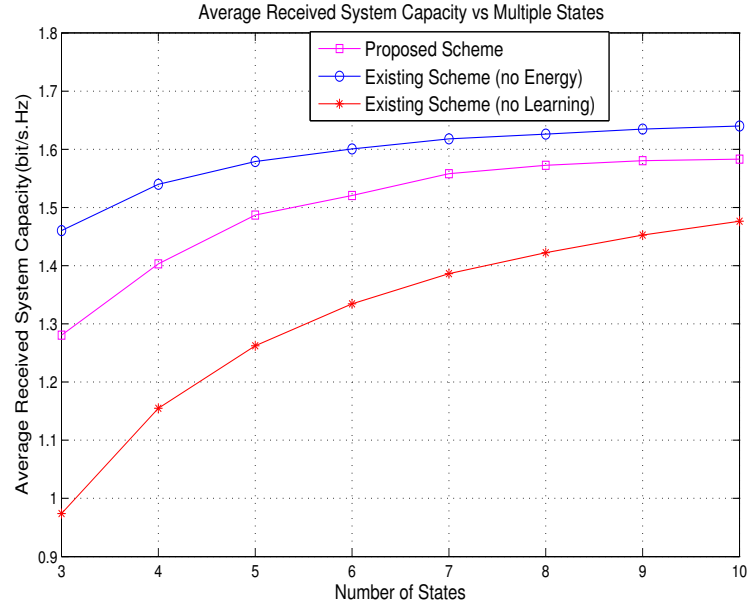


Fig. 3. The average system capacity versus the number of states.

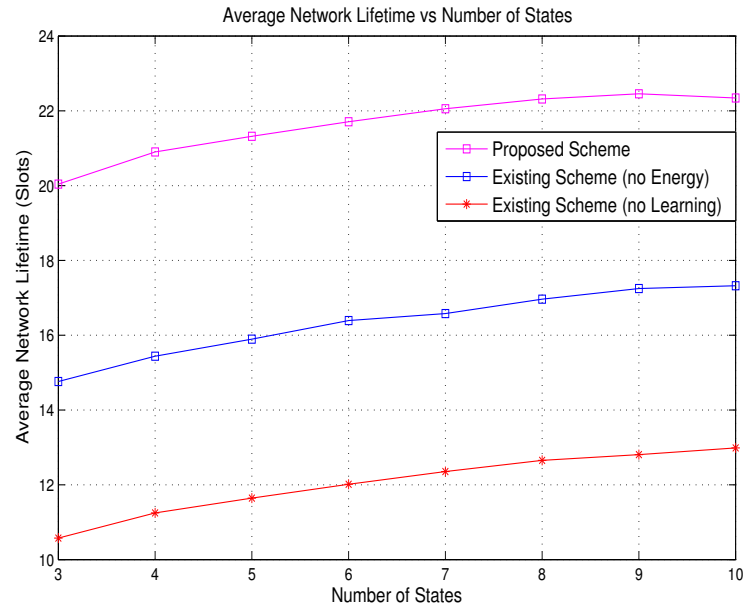


Fig. 4. The average network lifetime versus the number of states.

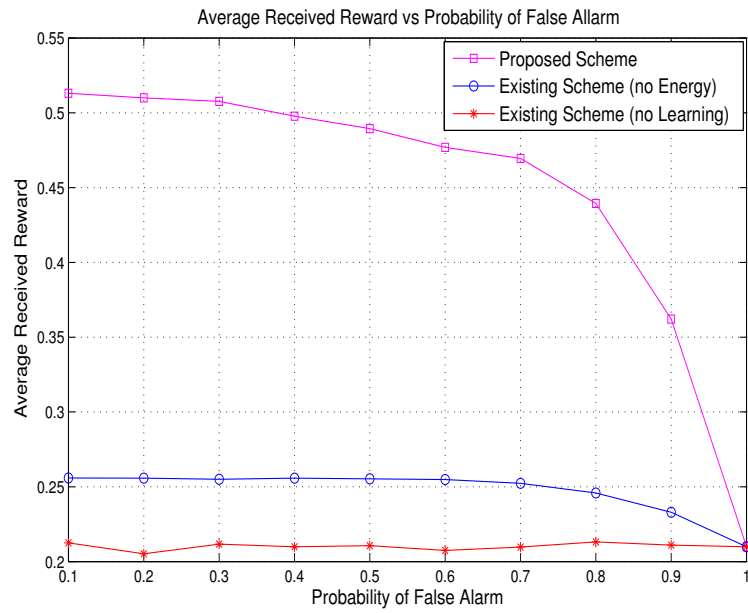


Fig. 5. The average received reward vresus the probability of false alarm.

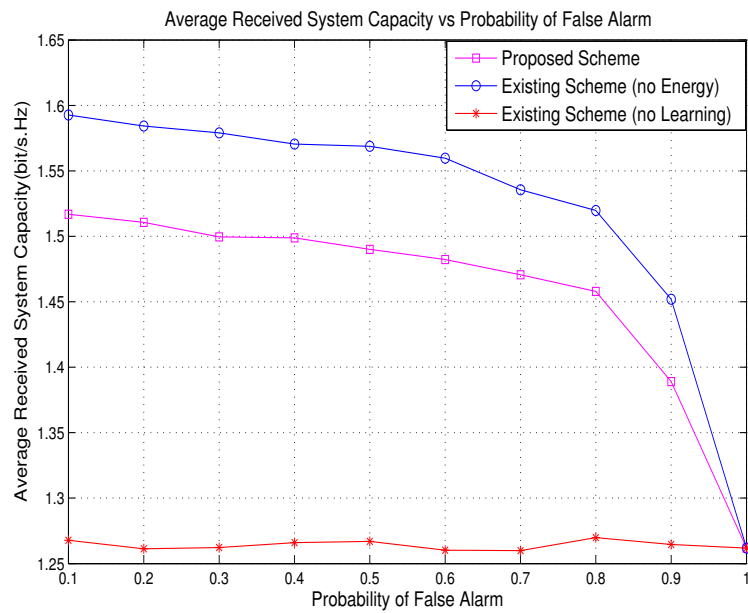


Fig. 6. The average received system capacity vresus the probability of false alarm.

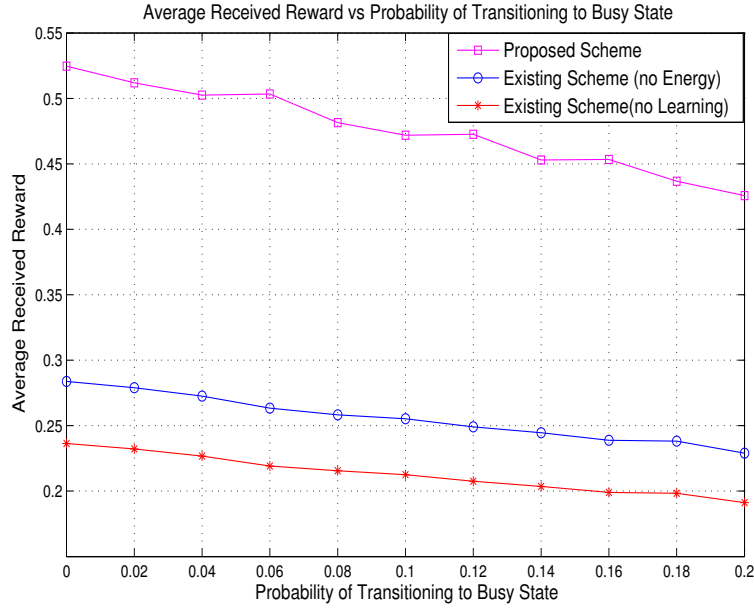


Fig. 7. The average received reward versus the probability of channel transitioning to the busy state.

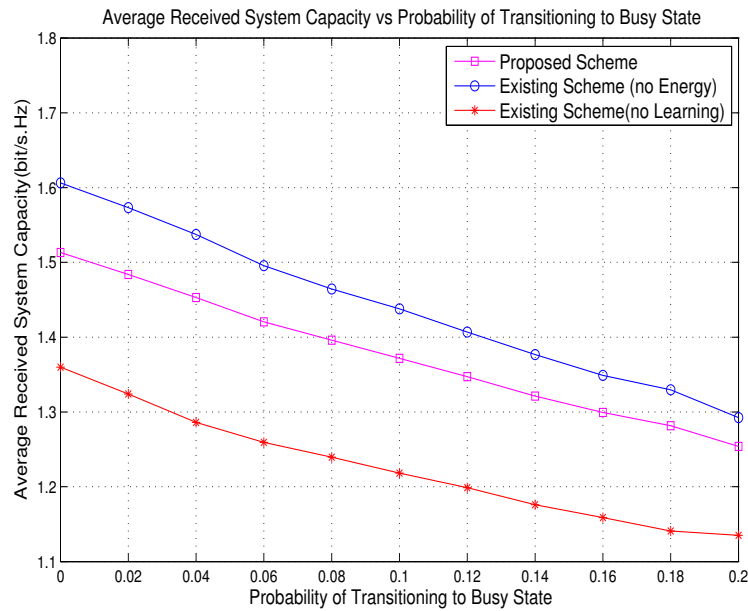


Fig. 8. The average received system capacity versus the probability of channel transitioning to the busy state.

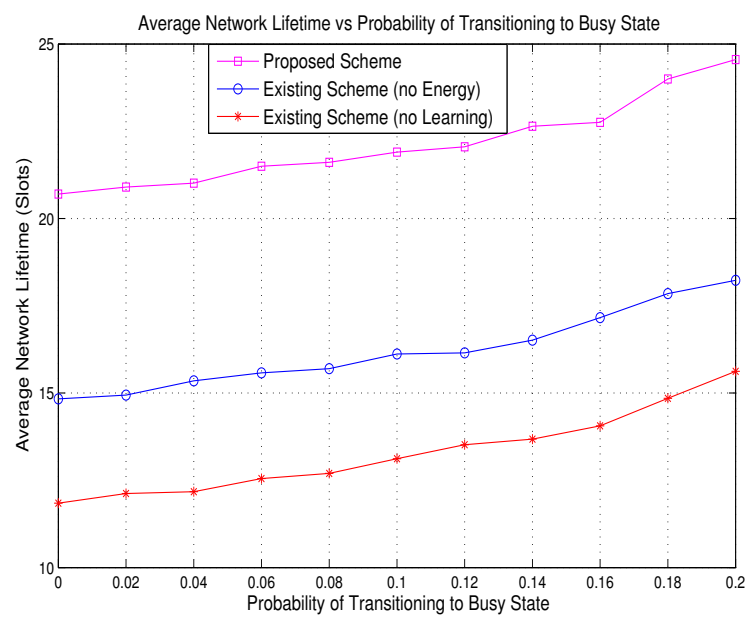


Fig. 9. The average Network Lifetime versus the probability of channel transitioning to the busy state.